

Standards in education

In everyday usage, the concept of standards is used to communicate either that something is good enough to be fit for purpose or that it has reached a particular level of excellence. Star systems for hotels tell us how good we can expect the accommodation to be.

British Safety Standards reassure us that child seats in cars can be trusted to work as we expect. In both cases the concept of standards embraces both a set of expectations (the straps will stay attached to the seat in case of collision; there will be TV and coffee facilities in the room...) and a means of assessing whether the expectations are met. This may involve laboratory tests, regular inspections following documented criteria or spot checks. The same dual concept exists with respect to standards in education.

In education, the concept of standards includes both the description of what is to be learned and taught (the curriculum) and the results of assessments. Educational assessments measure the extent to which learners have acquired knowledge, skills and understanding in a particular field, together with their ability to apply what they have learned. The outcome of an assessment may be reported as a mark or grade, or simply as pass or fail. When the outcome of an assessment is pass or fail, for example with a driving test, a pass result indicates that the learner has reached the minimum standard required to be regarded as competent but it does not indicate whether or not they have excelled as a driver.

When results are reported as grades, each grade indicates how well the learner has shown mastery of the required knowledge, skills and understanding and how well they have shown they can apply them, the higher grades denoting a higher standard. For assessments such as Cambridge IGCSE® and Cambridge International A Level, a particular grade gained in a particular subject indicates the same standard whether it is achieved through examinations taken in March, June or November, and regardless of the year in which they are taken. This enables us to talk in terms of a Grade A or a Grade C standard as a consistent level of achievement regardless of when or where the assessment was taken.

Maintaining the confidence of the wider education and employment communities in our ability to set a consistent standard in the awarding of grades is of key importance and Cambridge Assessment has a research programme dedicated to this work.



What terms are associated with standards in education?

While discussing assessment standards, which will be addressed in detail in subsequent sections, it is also useful to consider related key concepts.

Curriculum standards

These describe what learners are expected to learn at particular stages in their education. National curriculum standards appear in different forms. For example, the Common Core Standards in the US cover English and mathematics whereas the English national curriculum covers a wider range of subjects and has statutory status.

In order to know how far curriculum standards are being achieved, it is necessary to have assessments designed to measure this. In the case of New Zealand, the curriculum standards provide substantial detail about what will count as evidence that the standards are being met, to enable teachers to assess progress. In many countries the potential conflict of interest introduced when teachers are wholly or largely





responsible for assessing progress has led to the demand for, or existence of, external tests of one kind or another. In England, national tests at Key Stages 1 and 2 and GCSE examinations at Stage 4 are designed to measure progress and achievement against national curriculum standards.

International standards

There is no definitive set of international standards in education, either in terms of curriculum standards or in terms of assessment standards. However, the existence of, and rapidly growing interest in international surveys such as PISA, TIMSS and PIRLS has created and emphasised a notion of international standards: Those responsible for their country's education system frequently look on their ranking in international comparisons as a measure of how well their systems are performing.

Cambridge has designed curricula for use worldwide for ages 5–19, and our assessments at Cambridge Primary Checkpoint, Cambridge Secondary 1 Checkpoint, Cambridge IGCSE and Cambridge Advanced are designed to measure progress and achievement against these curricula. The emphasis of the Cambridge curriculum is on offering a consistent approach to learning while allowing the opportunity for school leaders to tailor their teaching to take account of local circumstances.

What is the theory behind assessment standards?

The key theoretical concepts that underpin assessment standards are validity, reliability and fairness.

In simple terms, an assessment meets the requirements of **validity** if it assesses what it says it does without requiring irrelevant knowledge or skills, and if what it

assesses really is the relevant set of knowledge and skills required by a contemporary study of the subject in question. For example, an assessment of Cambridge IGCSE mathematics should not require learners to have musical ability or knowledge of chemistry or the rules of football. It should require that they have studied the content set out in the Cambridge IGCSE mathematics syllabus, and the syllabus should include the mathematics that will be most useful to the learner and to their next teacher or employer. Validity requires that the inferences made by people using the outcomes of the assessment, for example that person X with a B grade is better prepared to make use of a university place than person Y with a D grade, turn out to be justified in practice. Newton and Shaw (2014) provides a comprehensive review of the concept of validity in assessment.

An assessment meets the requirements for **reliability** if the mark or grade reported as the result is correct. Reliability exists if repeat marking of a single assessment by a second or third examiner produces the same outcome, or if the learner achieves the same outcome when taking a second version of the same test. Maintaining the same standard for each grade from year to year is part of reliability.

There can be tension between validity and reliability. Multiple choice tests are the most reliable assessments because they allow no room for professional judgement in the marking. However, some things cannot readily be assessed by multiple choice, such as the ability to construct a proof or the ability to evaluate and draw conclusions. If these are the skills needed to meet the requirements of validity, then an assessment that uses only multiple choice questions cannot be valid, however reliable it is. On the other hand, if an assessment leaves so much room for professional judgement that there is little consensus between trained expert examiners about the correct mark it cannot meet the requirements of reliability and therefore cannot be considered valid.

Fairness in assessment requires both validity and reliability. It also requires that the conditions in which the tests are taken should be standard for all learners as far as possible. Consistency in the time allowed, the extent of use of reference materials and absence of assistance from others must be assured. But fairness also requires that access to a test is available to everyone who seeks to take it. This includes provision of special arrangements for learners with visual, hearing or other impairments that make it harder for them to take the test. In the USA it also requires that adequate allowance be made for learners whose first language is not English. These arrangements must be sufficient to overcome the impairment without conferring advantages on the learners in question.

The most influential statement of requirements for the validity, reliability and fairness of assessments in practice





is the *Standards for Educational and Psychological Testing*, produced jointly by the American Educational Research Association, the American Psychological Association and the National Council for Measurement in Education. Cambridge has formal processes to ensure the validity and reliability of its assessments, set out in *The Cambridge Approach* and in Cambridge International Examinations' *Code of Practice*.

What are the benefits of assessment standards?

Assessment standards are valuable to communicate what has been achieved at a particular stage of education and how it can be compared with other achievements. They provide a benchmark that does not depend on a particular teacher or on the eloquence of the writer of a reference. Within a school, they provide feedback to learners and teachers alike about what has been mastered well and what requires further attention. This feedback is a valuable source of improvement in teaching and learning for current learners and for the future.

When assessment comes at the end of a phase of education, certification of the standards tells a potential employer, admissions officer, or anyone else who looks at a certificate, what the owner of the certificate has demonstrated that they know and are able to do. Grades may also be used in school evaluation and for system level accountability. Without consistency of standards, such grades are meaningless.

¹ See Education Brief on International surveys: PISA, TIMSS, PIRLS

² See Education Brief on The Cambridge learner and teacher attributes

What are the challenges of assessment standards?

One of the key challenges is ensuring that the Grade A or C standard remains consistent, given that different learners have taken different examination papers. Two apparently straightforward approaches do not work:

The first would be to give the same percentage of learners a Grade A each year. This is known as norm referencing or maintaining consistency of outcomes. It would only be appropriate if you had other evidence that the learners were equally well prepared each time the assessment is conducted. For example, it may be true that 48% of learners pass their driving test at the first attempt nationally, but nobody would advocate passing 48% of first time takers in each test centre each day, without reference to the quality of their driving.

The second straightforward way would be to give Grade A to everyone who achieved at least the mark that was required for an A grade last year. This would work if there was independent evidence that there was no difference in the difficulty of the test between the two years. Experience shows that this is often not true. It is often the case that learners find particular questions on a test more or less difficult than the test developers anticipated. If the test is slightly easier than last year, it is fair to require a slightly higher mark for a Grade A than last year.

To meet the requirements of validity, reliability and fairness we need to be sure that the same standard of achievement will be awarded the same grade in different

years. The awarding of grades is the end point of a process that decides how much of any change in the distribution of marks from one year to the next is caused by a change in the difficulty of the test, and how much is caused by a change in the preparedness of the learners who have taken the test.

In some cases the development of a test includes the pre-testing and calibration of items (questions) on the test. This means that some or all of the items have been tried out on a sample of learners and their marks have been statistically analysed. By measuring how difficult items are before the test is put together, the difficulty of the test compared to previous tests can be calculated before it is taken. Little or no professional judgement is then needed in deciding the minimum mark required for each grade. This approach uses Item Response Theory or Rasch analysis and there is a huge literature. Baker (2001) provides a straightforward introduction.

In other cases, and for almost all Cambridge IGCSE and Cambridge Advanced assessments, we do not have an exact measure of the difficulty of the test before it is taken, because pre-testing is not practically possible, and decisions about the minimum mark for each grade require the use of both professional judgment and statistical and other evidence, including:

- archive scripts at the key grade boundary marks from previous sessions
- information about the size and composition (for example country and type of school attended) of the cohort of examinees
- teachers' forecast grades
- the distribution of marks (mean, standard deviation, cumulative percentage of examinees at each mark)
- 'putative' grade distributions (grade distributions reflecting data about the results of matched learners on tests taken at an earlier stage of education)
- experts' judgments about the quality of work in a sample of scripts with marks around where the minimum mark for the grade is expected to be found
- experts' judgments about the difficulty of the question paper
- any other evidence suggesting that the assessment had previously been severely or leniently graded and needs to be brought into line with other tests intended to carry the same standard.

Each of these sources of evidence taken alone could be misleading, but when taken together they enable us to maintain standards that embody consistency of achievement from year to year. Over a period of years, the valid content for an assessment will change, for example as a result of social changes and scientific discoveries. In response to these advances, Cambridge follows procedures to ensure that an

assessment remains fair by meeting the requirements for validity and reliability.

Recent educational debate has centred on how best to prepare learners for the challenges of the modern international world and the teaching and assessment of what are frequently referred to as '21st century skills'. Within this context, the effectiveness of current assessment practices as instruments to measure the standards of such skills has been questioned. In order to support schools in nurturing learners who can effectively participate in the modern world, Cambridge has developed the 'Cambridge learner' and 'Cambridge teacher' attributes. These refer to five highly desirable habits in learning representing a combination of values, attitudes, motivation, empathy, knowledge and skills. Cambridge subject curricula and assessment specifications are designed with these attributes in mind, but they need to be supported by teachers in the classroom.

Practical tips

Many countries have their own national curriculum standards but this does not guarantee that all or indeed any learners will achieve what the standards set out. Although the standards exist, the classroom reality may be very different. It is important to find out whether school leaders and teachers have access to the standards, and whether they are trained or have the necessary facilities in



schools to support teaching to the standards. In all cases national curriculum standards embody the aspirations of at least some of those who are charged with responsibility for education, and differences between national curricula reflect differences in culture and values.

The existence of the international surveys leads to a tendency to focus on rank orders produced by PISA, TIMSS and PIRLS. A frequent request when Cambridge works with a ministry of education on reform projects is that they want to retain their national curriculum at the same time as having assessments designed to measure progress against it, and they also want the assessments to be benchmarked against 'the international standard'.

We have used two approaches in responding to questions about how our curriculum and assessments can help school systems to meet international standards. We have mapped our curricula against other curricula – for example, against that of the country in question, and against the curricula of 'high performing jurisdictions', i.e. countries that have high ranking in the international surveys, and against the TIMSS Maths and Science curriculum. We have also arranged for learners to take selected Cambridge assessments or developed new assessments, and have used the outcomes to evaluate

their relative strengths and weaknesses.

How is Cambridge supporting schools with Standards?

The syllabus documents for Cambridge IGCSE and Cambridge Advanced provide clear information about the curriculum that will be assessed, and for Cambridge IGCSE and Cambridge Pre-U, descriptors of the kind of work that will reach the required standard for different grades. We publish question papers, mark schemes and a Principal Examiners' Report for Teachers for each subject after each administration of the examinations. For each subject there are also booklets of Example Candidate Responses that show how the descriptors work in practice.

Where can you find more information?

- Baker, F. (2001) *The Basics of Item Response Theory*. US: ERIC Clearinghouse on Assessment and Evaluation. echo.edres.org:8080/irt/baker
- *The Cambridge Approach: Principles for designing, administering and evaluating assessment* (2009) www.cambridgeassessment.org.uk/Images/109848-cambridge-approach.pdf
- *Cambridge International Examinations Code of Practice* (2008): www.cie.org.uk/images/7881-code-of-practice.pdf
- Core Curriculum Standards, United States of America: www.corestandards.org
- National Curriculum Standards, England: www.gov.uk/national-curriculum/overview
- National Curriculum Standards, New Zealand: nzcurriculum.tki.org.nz/National-Standards
- Newton, P. and Shaw, S. (2014) *Validity in Educational and Psychological Assessment*. London: Sage with Cambridge Assessment.
- Research Matters (2011) *Special Issue 2: Comparability*. UK: Cambridge Assessment.
- The Standards for Educational and Psychological Testing (2014). AERA, APA, NCME

Acknowledgement: Diane Palmer

Learn more! If you would like to know more about Cambridge Training please email info@cie.org.uk or visit www.cie.org.uk/events or contact Customer Services on +44 (0)1223 553554